

# Journal of Biomedical Optics

BiomedicalOptics.SPIEDigitalLibrary.org

## **Toward a functional near-infrared spectroscopy-based monitoring of pain assessment for nonverbal patients**

Raul Fernandez Rojas  
Xu Huang  
Keng-Liang Ou

**SPIE.**

Raul Fernandez Rojas, Xu Huang, Keng-Liang Ou, "Toward a functional near-infrared spectroscopy-based monitoring of pain assessment for nonverbal patients," *J. Biomed. Opt.* **22**(10), 106013 (2017), doi: 10.1117/1.JBO.22.10.106013.

# Toward a functional near-infrared spectroscopy-based monitoring of pain assessment for nonverbal patients

Raul Fernandez Rojas,<sup>a,\*</sup> Xu Huang,<sup>a</sup> and Keng-Liang Ou<sup>b,c,d</sup>

<sup>a</sup>University of Canberra, Human-Centred Technology Research Centre, ESTEM Faculty, Canberra, Australia

<sup>b</sup>Taipei Medical University Hospital, Department of Dentistry, Taipei, Taiwan

<sup>c</sup>Taipei Medical University-Shuang Ho Hospital, Department of Dentistry, New Taipei City, Taiwan

<sup>d</sup>3D Global Biotech Inc., New Taipei City, Taiwan

**Abstract.** Pain diagnosis for nonverbal patients represents a challenge in clinical settings. Neuroimaging methods, such as functional magnetic resonance imaging and functional near-infrared spectroscopy (fNIRS), have shown promising results to assess neuronal function in response to nociception and pain. Recent studies suggest that neuroimaging in conjunction with machine learning models can be used to predict different cognitive tasks. The aim of this study is to expand previous studies by exploring the classification of fNIRS signals (oxyhaemoglobin) according to temperature level (cold and hot) and corresponding pain intensity (low and high) using machine learning models. Toward this aim, we used the quantitative sensory testing to determine pain threshold and pain tolerance to cold and heat in 18 healthy subjects (three females), mean age  $\pm$  standard deviation ( $31.9 \pm 5.5$ ). The classification model is based on the bag-of-words approach, a histogram representation used in document classification based on the frequencies of extracted words and adapted for time series; two learning algorithms were used separately, K-nearest neighbor (K-NN) and support vector machines (SVM). A comparison between two sets of fNIRS channels was also made in the classification task, all 24 channels and 8 channels from the somatosensory region defined as our region of interest (RoI). The results showed that K-NN obtained slightly better results (92.08%) than SVM (91.25%) using the 24 channels; however, the performance slightly dropped using only channels from the RoI with K-NN (91.53%) and SVM (90.83%). These results indicate potential applications of fNIRS in the development of a physiologically based diagnosis of human pain that would benefit vulnerable patients who cannot self-report pain. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JBO.22.10.106013](https://doi.org/10.1117/1.JBO.22.10.106013)]

Keywords: pain; near-infrared spectroscopy; biomarker; machine learning; discrete wavelet transform; wavelet; principal component analysis; means; bag-of-words; K-nearest neighbor; support vector machine.

Paper 170430R received Jul. 17, 2017; accepted for publication Oct. 2, 2017; published online Oct. 27, 2017.

## 1 Introduction

Pain is a subjective experience, and no objective clinically available diagnostic test exists to measure it. However, in clinical practice, there are two main approaches for the assessment of pain: self-reported and clinical judgment. Self-reported methods try to rate patient's pain by verbal or numeric self-rating scales, such as visual analogue scales, verbal descriptor scales, numerical rating scales, or the MacGill pain questionnaire. Conversely, pain assessment by clinical judgment is based on testing and learning from observations of the type, significance, and context of the patient's pain perception. Self-report is the most accepted method and provides the most valid assessment in clinical practice. However, when self-reports are unavailable or in doubt, observational measures can be used as a complement or substitute.<sup>1</sup>

The absence of verbal (or writing) communication in some patients (also referred as nonverbal) is an obstacle to the evaluation of pain. Patients with impaired communication, unconscious patients, infants, the critically sick, persons

suffering from advanced dementia, and patients with intellectual disabilities are examples of vulnerable individuals, who are unable to speak for themselves.<sup>2</sup> Due to the inability to communicate pain status, these populations are at risk to be under- or overtreated while in pain. These conditions create a significant obstacle to evaluate and manage patient's pain experience in a suitable manner. Hence, the need for a reliable and objective pain assessment to assist medical practitioners in the diagnosis of pain is critical for this vulnerable population.

Pain assessment is not a trivial task and some factors should be taken into consideration to achieve a reliable estimate of it. For instance, pain perception can be affected by age, sex, weight, psychological state, or cultural background.<sup>3</sup> Similarly, Edwards et al.<sup>4</sup> found significant differences between African-Americans to report greater levels of pain than whites for conditions, such as AIDS, migraines, postoperative pain, joint pain, or arthritis; these findings were consistent across different age groups. In addition, tolerance to pain in patients differs due to the capacity of each person to resist (or adapt) different intensities of pain even within the same ethnic groups.<sup>5</sup> Thus, assessing and managing pain should be carried out to tailor individual needs.

Some strategies have been proposed to aid the objective assessment of pain. In clinical settings, for instance, the use

\*Address all correspondence to: Raul Fernandez Rojas, E-mail: [raul.fernandezrojas@canberra.edu.au](mailto:raul.fernandezrojas@canberra.edu.au)

of vital signs, metabolic markers, and brain imaging tools has been explored.<sup>6</sup> Vital signs, such as heart rate, blood pressure, respiratory rate, and arterial oxygen saturation, are commonly used by medical practitioners to measure pain in nonverbal patients. However, these parameters can be unreliable in situations, where the patient suffers from a mental disorder, is under stress, is premedicated, or is sedated. In these cases, changes in vital signs can be expected and relying on them could affect the interpretation and assessment of pain in such patients. On the other hand, the use of neuroimaging techniques [e.g., magnetic resonance imaging (MRI), positron-emission tomography (PET), electroencephalography (EEG), near-infrared spectroscopy (NIRS), etc.] has gained recognition due to their capabilities to gain insights into the human brain (noninvasively) and understand the components involved in pain processing.

In that context, functional near-infrared spectroscopy (fNIRS) can be used to determine hemodynamic and metabolic changes associated with brain activity by measuring changes in the concentration of oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (HbR) in real time. This neuroimaging technique can be used to assess cortical activity in diverse experimental and clinical settings, such as neurology (e.g., Alzheimer's disease, dementia, stroke recovery), psychiatry (e.g., anxiety, schizophrenia, personality disorders), psychology (e.g., attention, language, emotion), and other applications (e.g., brain-computer interface, sports sciences, pain research).<sup>7</sup> By placing the fNIRS optodes on specific cortical regions, it is possible to access the hemodynamic response associated with neural activity. The activation of cortical regions is the result of vascular dilation increasing cerebral blood flow, thus, fluctuations in the hemodynamic response correlate with neural activity.<sup>8</sup> In addition, fNIRS offers advantages over other technologies (fMRI, EEG, PET) such as it is noninvasive, portable, relatively inexpensive, and provides good temporal and spatial resolution,<sup>9-11</sup> which makes this technology ideal for clinical testing.

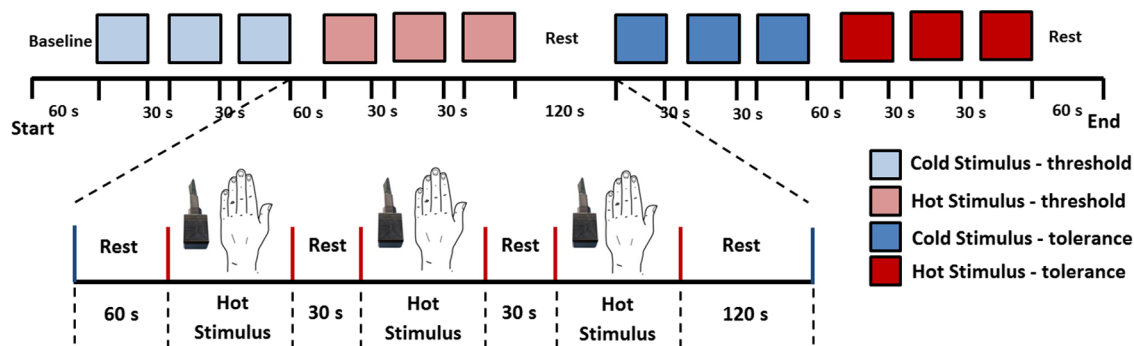
In the field of pain research, fNIRS studies have demonstrated that external pain stimuli in healthy and diseased patients evoke changes of oxygenation levels in distinctive cortical regions.<sup>6,12-14</sup> Several techniques have been utilized to safely evoke noxious stimulation, including mechanical, thermal, electrical, and chemical.<sup>8</sup> For example, Yücel et al.<sup>8</sup> used innocuous and noxious electrical stimuli in 11 healthy subjects and obtained a localized hemodynamic response in the right and left motor-sensory regions; the noxious stimuli resulted in a higher response (HbO) as compared to the innocuous stimuli. In another study, Lee et al.<sup>15</sup> observed distinct activation in the parietal and frontal areas for acute pressure pain and histamine-induced itch in seven subjects; the results showed a faster activation and stronger response of HbO signals for pressure pain than itch stimulation in both the frontal and parietal area. Ranger et al.<sup>16</sup> recorded the HbO response of 20 critically ill infants (<12 months of age) during noxious procedures for cardiopulmonary bypass for various congenital heart defects (CHD) and found a significant bilateral increase in the somatosensory region despite the administration of analgesic treatment. In another study, Slater et al.<sup>17</sup> studied changes of total hemoglobin (HbT = HbO + Hb) in response to noxious stimulation (heel lance) in 18 infants aged between 25 and 45 weeks (post-menstrual age) over the somatosensory cortex; the noxious stimulation produced a clear hemodynamic response in the

contralateral somatosensory cortex and was significantly greater in awake compared to sleeping infants. These studies have showed the potential to use fNIRS to obtain pain-related information in human subjects.

Similarly, attempts to use neuroimaging methods to identify pain in healthy humans have shown potential to aid the assessment of pain. For example, Brown et al.<sup>18</sup> used fMRI data from eight individuals to train a support vector machine (SVM) to distinguish painful from nonpainful thermal stimuli with 81% accuracy. In another fMRI study, Wager et al.<sup>19</sup> predicted pain intensity and identified patterns of noxious heat from non-painful warmth using a principal components (PCs) regression technique with a performance of 93%. Gram et al.<sup>20</sup> used EEG to investigate morphine- and placebo-administered subjects after receiving stimulation using the cold pressor test, the authors used SVM to classify responders with an accuracy of 72%. In an fNIRS study, Pourshoghi et al.<sup>21</sup> used a SVM classifier to classify low pain and high pain from healthy subjects after a cold pressor test with 94% accuracy. These results show that pain recognition and classification is plausible using neuroimaging methods, also the results from these studies advocate for the use of machine learning techniques to predict human pain.

Significant progress on the work of the aforementioned studies would be to design classification models that can discriminate multiple pain signatures at different intensities. For example, using two types of thermal stimuli (hot and cold) and different intensities (low and high), the learning model would be more robust than a model for a single pain condition (cold or heat). Another contribution would be to show the ability of bag-of-words (BoW) to represent time series in the field of brain imaging, which in conjunction with machine learning are a promising tool in neuroscience research. Therefore, these new classification models, based on robust learning models capable of differentiating multiple signatures of pain at different intensities and which could be extended to nonverbal patients, would be more valuable for clinical testing.

In this study, we aim to contribute toward the establishment of a tool for pain assessment in nonverbal patients using fNIRS and machine learning techniques. To that end, we applied the threshold test and tolerance test as measured in the quantitative sensory testing (QST) to obtain the thermal sensory profile in 18 healthy subjects using cold and heat. Therefore, we explored the possibility to classify fNIRS signals of pain according to temperature level (cold or hot) and pain intensity (low or high). Two machine learning classifiers [K-nearest neighbor (K-NN) and SVM] were trained to classify fNIRS between cold and heat pain while distinguished between low and high pain intensity; we built our classifiers using the BoW representation used in text classification. The major contributions of this study are (1) heat and cold pain are differentiated using fNIRS signals (HbO) as well as their corresponding pain intensity (low or high) using machine learning techniques; (2) we report the investigation of two state-of-the-art classifiers to distinguish between thermal noxious stimuli; and (3) we also investigated the accuracy of our classifiers with respect to the region of interest (8 channels) or whole head probe (24 channels). This study represents a step closer to developing a physiologically based diagnosis of human pain that would benefit vulnerable patients, who cannot self-report pain.



**Fig. 1** Stimulation paradigm. Pain threshold test was first measured and followed by pain tolerance test. In each test, cold and heat stimulus were applied on the back of the hand of each subject.

## 2 Methods

### 2.1 Subjects

Eighteen right-handed volunteers (three females) were considered in the study. All participants were of Chinese ethnicity and dark brown hair color, mean age  $\pm$  standard deviation ( $31.9 \pm 5.5$ ). No participants reported a prior history of neurological or psychiatric disorder, a current unstable medical condition, or under medication at the time of testing. All participants were right-handed to avoid any variation in functional response due to lateralization of brain function. The experimental procedure was explained to the participants, and they had the opportunity to stop the procedure at any time if it was needed. Written informed consent was obtained from all volunteers after providing an explanation of the protocols in the study. Procedures and methods for this study followed the guidelines accepted by the Declaration of Helsinki. This research study was approved by the Taipei Medical University (TMU) and full-board review process of the TMU-Joint Institutional Review Board under contract number 201307010.

### 2.2 Experimental Paradigms

Thermal pain perceptions (threshold and tolerance) in healthy volunteers were investigated following the standard procedures of the QST protocol.<sup>22</sup> We defined pain threshold (low pain) as the least stimulus intensity at which stimulation becomes painful and pain tolerance (high pain) as the highest intensity of pain at which stimulus becomes unbearable. Both thermal threshold and thermal tolerance were obtained by heat and cold stimulation, and pain measurements were obtained on the dorsum of the left hand (nondominant hand). All thermal tests were performed using a pathway CHEPS (Medoc, Israel), this sensory analyzer delivers heat and cold to the skin with a thermode. Attached to the central unit is a hand-held controller that allows subjects to press a button to stop the heating/cooling of the thermode. The subjects were seated on a comfortable armchair with the left arm resting on a table. Previous studies have reported no significant difference in QST between the right and left sides of the body.<sup>22</sup>

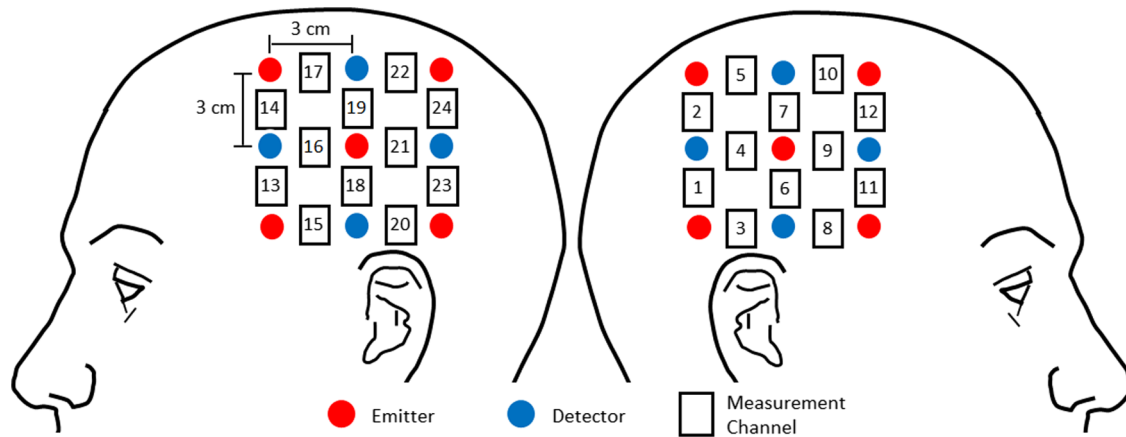
The pain experiment involved applying cold and heat to the skin to induce pain. In this test, subjects are exposed to gradually increasing or decreasing temperatures with a thermode and they pressed a button when they experienced pain (threshold test) and highest intensity of pain (tolerance test). The temperature of the thermode, just as it becomes painful or unbearable, is the thermal pain threshold or thermal pain tolerance, respectively.

The pain threshold and tolerance for each subject were estimated as the average from three consecutive trials of cold and heat stimulation. Figure 1 summarizes the stimulation paradigm and is explained below.

Similar to Rolke et al.,<sup>22</sup> our protocol was divided in two tests: the thermal pain threshold (low pain) and the thermal pain tolerance (high pain), with a 2-min rest between both tests. After an initial 60-s rest, the thermal pain threshold was first measured by three consecutive trials of cold stimulation, followed by 60-s rest and then, three consecutive measurements of heat stimulation. Thermal pain tolerance was measured by following the same procedure for thermal pain threshold. In both tests, there was a 30-s interval between each trial. The three consecutive measurements from each test were averaged, and the block-averaged signal ( $\sim 30$  s) from each thermal test was used for the classification task. Thermal measurements were obtained with ramped stimuli of  $1^\circ\text{C}/\text{s}$  that were terminated when the subjects pressed a button, then, the temperature returned to baseline ( $32^\circ\text{C}$ ) for the next stimulation with a rate of  $10^\circ\text{C}/\text{s}$ . The probe has a safety cut-off to prevent any harm to the skin. The low cut-off temperature was  $0^\circ\text{C}$  and the high cut-off temperature was  $50^\circ\text{C}$ . The contact area of the thermode was  $9\text{ cm}^2$ . The testing took place in a quiet and temperature-controlled ( $20^\circ\text{C}$  to  $22^\circ\text{C}$ ) room at TMU. All experiments were carried out by the same researcher to reduce variabilities between samples. Based on the threshold and tolerance of cold and heat stimuli, the fNIRS data were organized into four categories: (1) low-cold (low pain), (2) low-heat (low pain), (3) high-cold (high pain), and (4) high-heat (high pain). These categories (1 to 4) were used to label the database and employed as classes for the classification task.

### 2.3 fNIRS Recording, Preprocessing, and Channel Selection

fNIRS recordings were acquired using a multichannel optical topography system, Hitachi ETG-4000 (Hitachi Medical Corporation, Japan). This equipment uses two wavelengths of near-infrared light to measure changes of hemoglobin concentration in the cerebral cortex; HbO at  $695\text{ nm}$  and HbR at  $830\text{ nm}$ .<sup>9</sup> The spectrometer is equipped with a 24-channel cap, configured in 12 channels (rectangles) per hemisphere (Fig. 2). Each hemispheric probe consists of five sources (red circles) and four detectors (blue circles) that provide 12 source–detector pairs. The head probe has a predefined source–detector separation (SDS) of  $3\text{ cm}$  (left panel Fig. 2) and an NIR-light head



**Fig. 2** Channel location and configuration. Channel probes were located around the C3 and C4 areas. For classification purposes, two sets of channels were tested, all 24 channels and 8 channels from a RoI. The selected channels from the RoI are four channels on the right hemisphere (Ch4, Ch6, Ch7, Ch9) and four channels on the left hemisphere (Ch16, Ch18, Ch19, Ch21), which are positioned over the primary somatosensory region. Source–detector distance was 3 cm.

penetration of about half the SDS distance.<sup>23</sup> According to the EEG 10 to 20 system, the measuring probes were centered on the C3 and C4 positions. In this study, we used only the HbO signals because HbO signals exhibit a better signal-to-noise ratio than HbR signals.<sup>24,25</sup> The sampling rate was 10 Hz.

Raw fNIRS data are normally contaminated by different sources of noise, and preprocessing was required. To minimize effects of physiological components (e.g., cardiac, respiration) and electrical noise in fNIRS data, a finite-impulse fourth-order low-pass filter with a cut-off frequency of 0.1 Hz was used. Then, motion artifacts were removed by following a denoising procedure similar to Molavi et al.,<sup>26</sup> which uses discrete wavelet coefficients to find and eliminate those coefficients (outliers) that do not belong to the probability distribution and therefore behave as motion artifacts; four subjects presented motion artifacts. Raw signals were converted to changes in HbO and HbR concentrations using the modified Beer–Lambert law applying literature differential path length factor values of 6.51 for the 695 nm and 5.86 for the 830 nm light sources.<sup>27</sup> In order to reduce extracerebral hemodynamics (scalp) and systemic variables<sup>23</sup> (e.g., blood pressure, breathing rate, autonomic nervous system activity) existed in our data, we followed the method presented by Zhang et al.<sup>28</sup> we applied principal component analysis (PCA) to the baseline data, assuming that baseline data primarily contain the spatial patterns of systemic interference and identified the first PC (which accounted for approximately 90% of the variance); then, we attenuated the interference in the stimulus data by projecting the stimulus data onto the orthogonal subspace of identified spatial eigenvectors to obtain “clean” stimulus data. Finally, we normalized all fNIRS data to zero mean and unit standard deviation to avoid different baselines and amplitude variations.

The number of fNIRS channels used in this study was evaluated during the classification task and two sets of channels were selected for that purpose. The first set of channels is formed of all the 24 channels available, whereas the second set is focused on a specific region-of-interest (RoI) composed of eight channels. The reasoning behind this evaluation is that the fNIRS head probe was covering a wide area of the cerebral cortex, including the frontal and parietal lobes on both hemispheres. However, for this particular study, we were interested in evoked signals related

to pain in the left hand, therefore, our RoI was the somatosensory region in the parietal lobe near the central sulcus on both hemispheres. Based on this hypothesis, we decided to use the channels around this cortical area and compare the performance of RoI-based channels and all the 24 channels. The RoI-based channels from both hemispheres were: Ch4, Ch6, Ch7, Ch9, Ch16, Ch18, Ch19, Ch21.

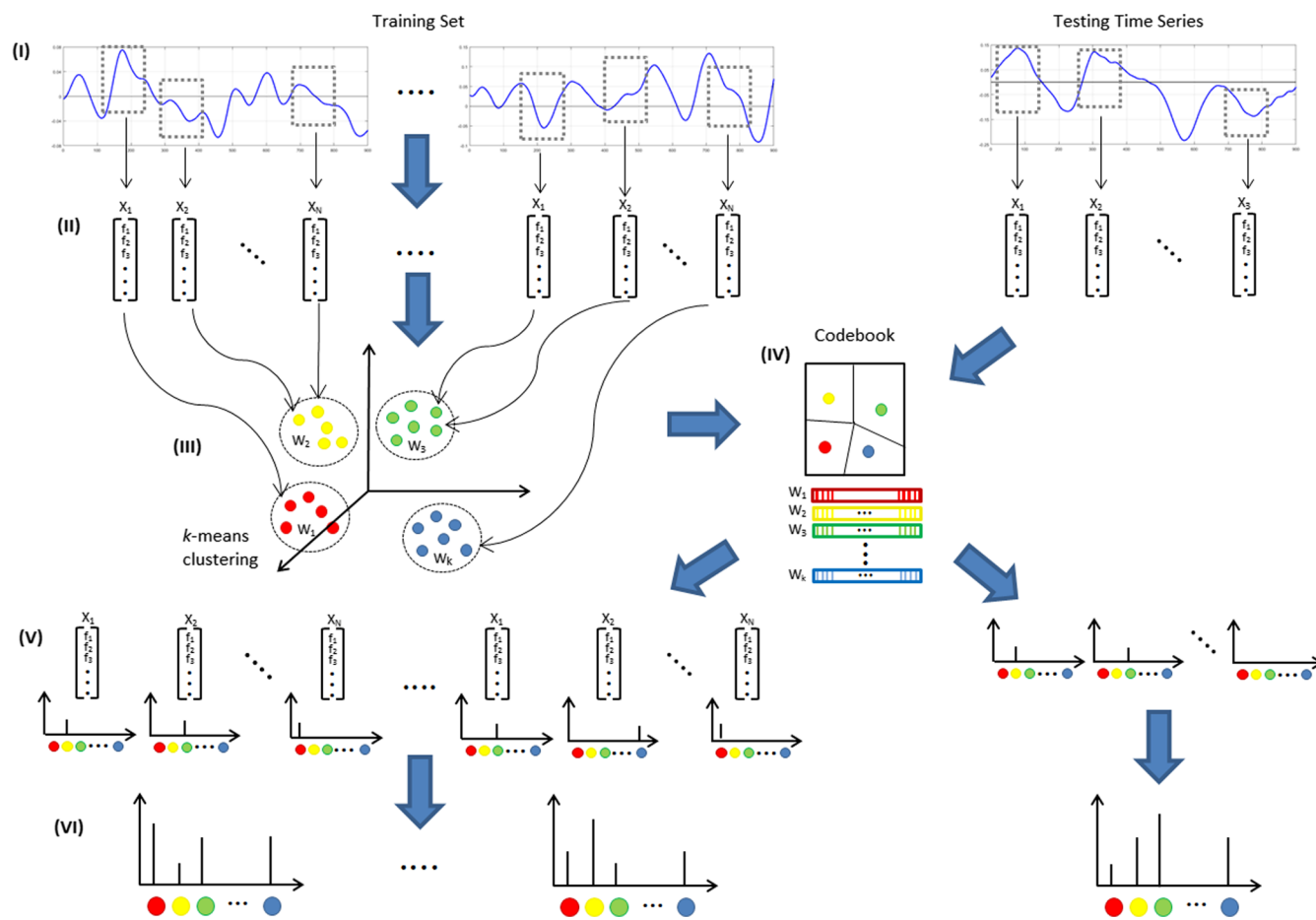
## 2.4 Bag-of-Words Representation

This study is based on the BoW approach for document classification adapted for time series. BoW representations are very popular in text mining, document classification, texture analysis, scene classification, and robot localization due to its simplicity and good performance.<sup>29,30</sup> BoW has become well-established in the field of computer vision, where it is also referred to as bag-of-features. Recently, it has been adapted for the classification of time series in audio and speech recognition,<sup>31–33</sup> electroencephalogram (EEG) and electrocardiogram (ECG) signals,<sup>34</sup> and time series similarity.<sup>35</sup> In these works, time series are treated as text documents and sections extracted from the time series as words.

In the context of fNIRS time series representation, BoW can adopt a similar model to the approach used for image classification. It can be generalized in three steps: first, the detection of keypoints (features) in the fNIRS time series using sliding windows; second, keypoints are represented into words (codebook generation) using a quantization technique (e.g., *k*-means); and third, fNIRS time series are characterized by histograms of number of occurrences for each word observed in each time series. This histogram representation is used as input to a classifier (e.g., SVM) that finds boundaries between classes. These steps are exhibited in Fig. 3 and are described below.

### 2.4.1 Feature extraction

The method starts by sliding a window through the fNIRS (HbO) time series and obtaining local features from each partition. Consecutive windows are obtained with incremental steps to divide each fNIRS time series. Local features are obtained from every sliding window and represented by their corresponding feature vector. The influence of different step size and window length was evaluated during the training process.



**Fig. 3** fNIRS time series representation: (I) sliding a window over the entire time series, (II) local segments are converted to feature vectors, local descriptors (e.g., wavelet) can be used to represent feature vectors, (III) vector quantization using *k*-means clustering, (IV) codebook (or dictionary) generation, each cluster center produced by *k*-means becomes a word ( $W_k$ ), the codebook is universal for training and testing data, (V) each feature vector is assigned to nearest data in the dictionary, (VI) histogram representation containing total number of appearances of each word in the time series. For new time series (testing data), repeat steps (I), (II), (V), and (VI).

In the literature, different methods for feature extraction have been employed to describe time series into BoW. Wang et al.<sup>34</sup> used discrete wavelet transform (DWT) from sliding windows to obtain a number of wavelet coefficients using a Daubechies order 3 (db3) wavelet transform. Grabocka et al.,<sup>36</sup> extracted coefficients of a polynomial regression by a sliding window, and the degree of the polynomial was selected based on the classification accuracy over a validation test. Baydogan et al.,<sup>30</sup> composed each feature vector by the slope of the fitted regression line, mean and variance of the values, and start and end points of the experiment. In our study, we used wavelet db8 approximation coefficients because we were interested in the wavelet coefficients from the period of stimulation (~30 s) in the activation band (0.02 to 0.08 Hz).<sup>37</sup> DWT has shown better results compared to other methods for the classification of time series.<sup>34</sup> In addition, wavelet analysis has the ability to localize frequency content that mostly contributes to the total energy contained within the fNIRS signal at the specific stimulation period, and the approximation coefficients are more robust to noise than raw fNIRS signals (using a frequency band less contaminated by physiological signals) while downsampling by a factor of 2 the original signal.<sup>34</sup>

### 2.4.2 Codebook formulation

The next step is clustering of extracted feature vectors and the creation of the codebook. After the local features are extracted, features are quantized using the *k*-means algorithm and a codebook of *k* words is obtained. The *k*-means algorithm generates a *k* set of centroids from a set of feature vectors, and each feature vector is classified with the centroid index closest to it. Following this concept, all feature vectors gathered from the sliding window process on the fNIRS time series (training data) are clustered to their nearest centroid (cluster center). These centroids are referred to as the codebook (or word dictionary) and each centroid can be seen as a word in the dictionary. We can perceive each cluster as a group of feature vectors who share local pattern similarities and the centroid of each cluster as the most representative value of that specific local pattern.

The total number (*k*) of centroids is referred as the codebook size and it is regulated by the number of clusters formed in the clustering process. If a small codebook is chosen, the number of centroids under-represents all feature vectors producing a poor discriminative performance since two feature vectors could fall into the same cluster even if the vectors are different. On the

other hand, if a large codebook is obtained, it is more prone to noise due to the lack of generalization of the codebook while significantly increasing the computational complexity. In order to create a more discriminative and adequate codebook, a trade-off between discrimination and generalization should be considered.<sup>38</sup> Thus, the effect of different codebook size was also investigated.

### 2.4.3 Codebook assignment

The last step in the BoW representation is the assignment of all feature vectors to their closest centroid. From the previous steps, for a fNIRS time series, we slid a window one step at a time and, from each window, we obtained a set of local features. We then constructed a codebook from all local features using the  $k$ -means algorithm. Now, we represent each fNIRS time series as a histogram of  $k$  number of bins, where each bin symbolizes each word in the codebook and the magnitude of each bin is the number of occurrences for that specific word in the time series. Each histogram can be seen as a normalized representation of pattern occurrences (i.e., words) from the total number of patterns (codebook) in the whole database. It is worth noting that the temporal ordering of the local features in each time series is ignored, and the centroids indexing (i.e.,  $1, 2, \dots, k$ ) obtained in the clustering task is used to order each word within the histogram. The BoW representation of  $N$  number of fNIRS time series is referred as the  $N$  number of histograms gathered following this model. Therefore, the histograms are the new representation of the fNIRS time series. The codebook generation is only computed once from training data and is generic to be used on test data. These histograms are then passed to a classifier to learn how to map classes from this BoW representation.

## 2.5 Classifiers

The classification task is to assign a vector  $X_i$  of input  $\mathbf{X} = [X_1, X_2, \dots, X_i]$  to one of the  $j$  classes  $Y_j$ . For that purpose, a classifier has to learn a model from training data  $\mathbf{X}_{\text{train}}$  (a dataset reserved for this purpose). We then use this model to make class assignments on a different dataset  $\mathbf{X}_{\text{test}}$  (independent from training data). This model captures the relationship between features and class labels in the training data. The end goal of a classification problem is to make generalized predictions on unlabeled data from the same distribution using the trained model. Properly, a classifier can predict the label  $Y$  of a given feature vector  $X$  as a function  $f$  of the type  $Y = f(X)$ .<sup>39</sup> To obtain the best model on our pain database, two state-of-the-art algorithms were compared, the  $K$ -nearest neighbor (K-NN) and SVMs for multiclass classification (one-versus-one voting).

The  $K$ -nearest neighbor (K-NN) is one of the simplest classification algorithms for supervised learning and a robust classifier for time-series classification.<sup>40</sup> It uses the  $K$  features in the training set  $\mathbf{X}_{\text{train}}$  closest (nearest) in feature space to unlabeled data  $x_0$  to predict its output  $y_0$ . The algorithm basically uses a majority vote between the  $K$  nearest training points and the query point. Similarity is defined based on a distance metric between two data points, for that purpose, we used Euclidean distance due to its simplicity and low computational cost.<sup>41</sup> Different number of nearest neighbors ( $K = 1, 3, 5, 7, 9$ ) was also tested.

SVMs have been previously used in classification problems in the neuroscience field.<sup>18,21</sup> Its basic idea is to find a linear

separating boundary with the maximal margin, and new data points are then classified according to which side of the decision boundary they fall.<sup>42</sup> However, data in real-world scenarios are not easy to separate by a linear decision boundary; in these cases, SVM solves this problem by kernel functions to map the observations into higher-dimensional space in which the data can be separable. We used the LibSVM library<sup>43</sup> parametrized for a multiclass classification and the linear kernel was used for this study. One of the reasons to use a linear kernel compared to a higher-order kernel (e.g., radial base kernel) is that a linear kernel is faster, has less parameters to optimize, and it has been used in previous studies on pain research using brain imaging techniques.<sup>18</sup>

## 2.6 Evaluation of Classification Models

The pain database was randomly split (on the subject level) in two parts, training and testing datasets. The training set was formed with 13 subjects and the testing set with the remaining 5 subjects. Therefore, the data are divided into two independent sets, one subset is used to train the classifier and another subset is used to test the performance of the classifier. During the training process, leave-one-out cross validation on the subject level is computed. One data point (1 subject) is held as the validation set while the remaining data (12 subjects) are used for training; the process is repeated 12 times, validating on a different subject in each iteration. During this process, the BoW parameters (step size, window length, and codebook size) were evaluated with a 1-NN and a linear SVM. The trained models were then tested using the remaining five subjects and the estimates (accuracy, sensitivity, and specificity) of predictive performance are then averaged across all the testing groups and reported as the final result.

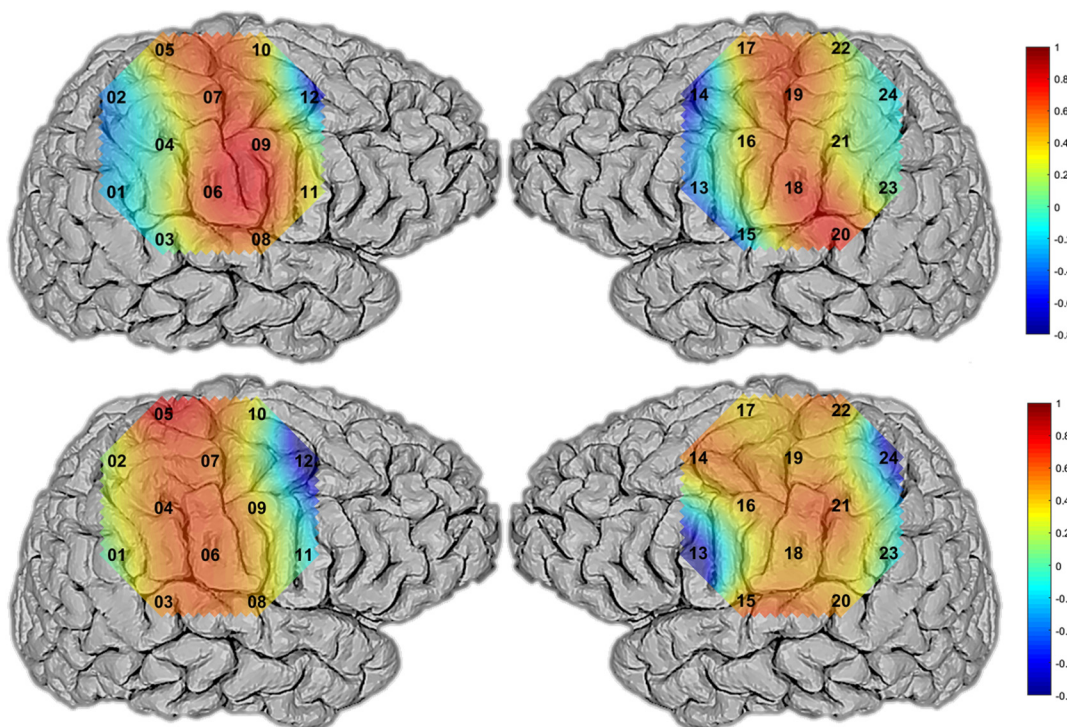
## 3 Results

Figure 4 presents an example of the activation map during stimulus. The image shows the hemodynamic response of all the 24 channels and the RoI. In addition, we can see that the activation is concentrated in the somatosensory strip. Therefore, we hypothesized that using relevant channels for the classification process would improve the performance of the classifiers.

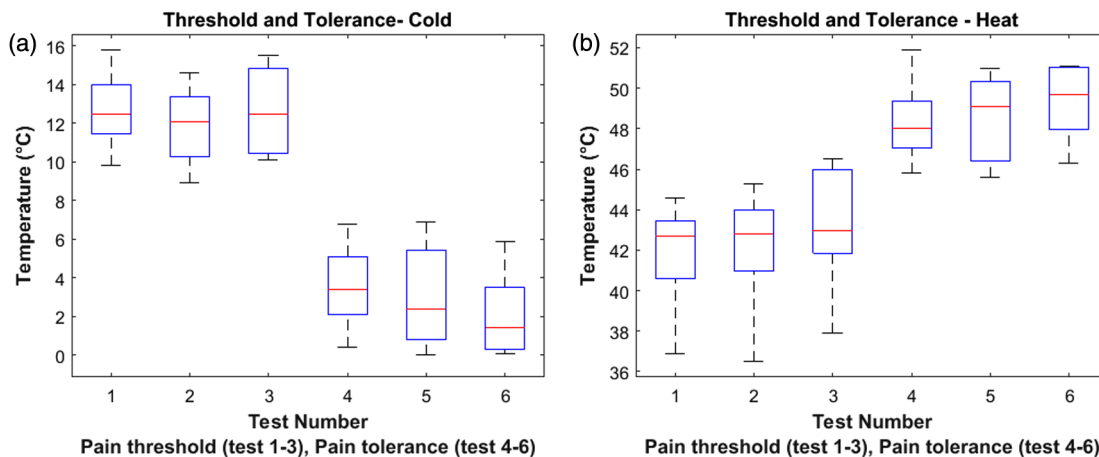
### 3.1 Thermal Threshold and Tolerance

Thermal threshold and tolerance of pain perception were obtained following a similar approach to the thermal test in the QST. By determining the QST thermal tests, we aimed to minimize the subjective nature of self-reported pain scores. In this way, the obtained measurements are based on temperature readings to classify a measurement as low pain (threshold) or high pain (tolerance) of cold and heat stimulation and not on self-reports. The measured (averaged) values obtained from each experiment are shown in Fig. 5.

The two plots show the threshold and tolerance temperatures of cold (Fig. 5, left panel) and heat (Fig. 5, right panel) stimuli across all participants. The first three measurements in each plot refer to the pain threshold while the last three measurements refer to the pain tolerance. The median ( $\pm$ SD) temperatures (horizontal red lines in Fig. 5) in which participants first perceived pain (pain threshold) from cold ( $12.45 \pm 1.97$ ,  $12.05 \pm 1.93$ ,  $12.45 \pm 2.22^\circ\text{C}$ ) and heat ( $42.70 \pm 2.44$ ,  $42.80 \pm 2.75$ ,  $42.95 \pm 2.92^\circ\text{C}$ ) clearly presented differences from the



**Fig. 4.** Stimulus-evoked hemodynamic activation (HbO), most of the activation is concentrated in the somatosensory region.



**Fig. 5** Thermal threshold and tolerance levels experienced by the participants (a) after cold and (b) heat stimuli. Horizontal red lines are the median values across all participants for each test.

highest intensity of pain (pain tolerance) the participants could take from cold ( $3.40 \pm 2.07$ ,  $2.40 \pm 2.71$ ,  $1.45 \pm 2.12^\circ\text{C}$ ) and heat ( $48.00 \pm 1.92$ ,  $49.10 \pm 2.11$ ,  $49.70 \pm 1.82^\circ\text{C}$ ). Median values (horizontal red lines) of cold pain threshold exhibited a statistically significant ( $p < 0.05$ ) higher response (Wilcoxon signed rank test,  $p = 0.0018$ ) than the median response to cold pain tolerance; similarly, median values of hot pain threshold showed a significant higher response (Wilcoxon signed rank test,  $p = 0.002$ ) to the average response to hot pain tolerance.

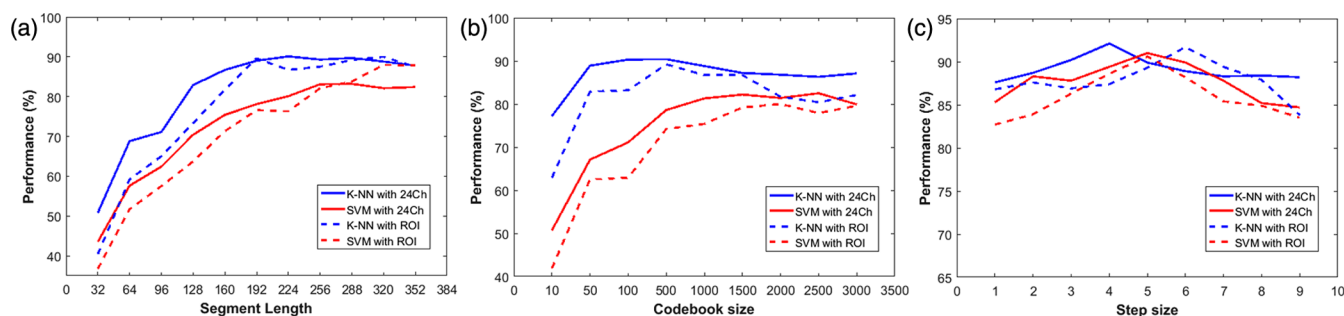
Habituation of noxious thermal stimulation was observed in all four experiments. As we can see in Fig. 5, group temperatures of pain threshold and pain tolerance slightly decreased and increased after each cold and heat stimulation, respectively. Therefore, we used Friedman’s test and looked for statistically

significant differences between the three consecutive measurements (test 1, test 2, test 3) from each thermal test (e.g., cold threshold). However, the Friedman’s test did not show any statistically significant differences ( $p < 0.05$ ) between repetitions within cold threshold ( $p = 0.18$ ), heat threshold ( $p = 0.57$ ), cold tolerance ( $p = 0.22$ ), and heat tolerance ( $p = 0.13$ ) tests.

### 3.2 Bag-of-Words Parameters

As a validity check on the effectiveness of the chosen parameters, the main parameters to build the BoW implementation were examined. Different step sizes, length of local segments, and codebook sizes were evaluated during the training process. The parameter evaluation included two classifiers (1-NN and





**Fig. 6** BoW parameter tuning. (a) Segment (window) length, (b) codebook size, and (c) step size were evaluated; their corresponding classification performance while testing the parameters is also presented. Note that the units of segment length and step size are samples, and the units of codebook size are words.

linear SVM) and two channel sets (24 channels and 8 channels). The results of parameter tuning of the BoW representation are presented in Fig. 6 and are described below.

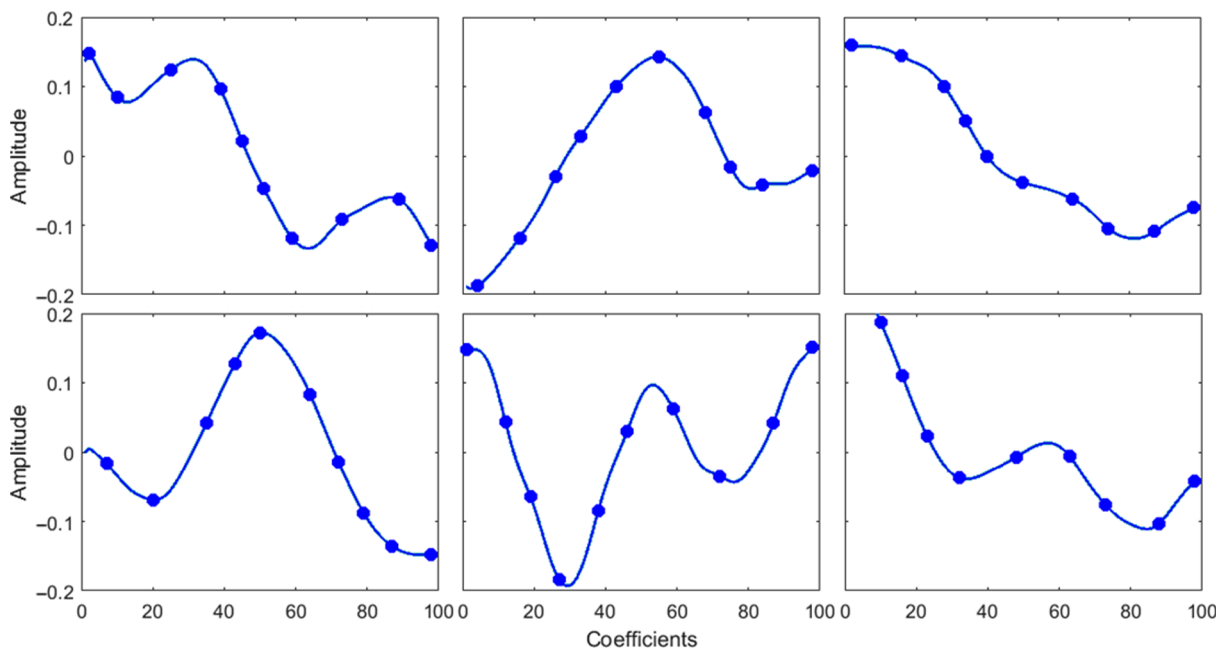
We first evaluated the length of the sliding window by varying its size among the range of 32 to 352 in increments of 32 samples. In this process, we fixed the codebook size to 1000 and the step size to 4; these reference values have shown good results in similar studies.<sup>30,34</sup> The best results were obtained using a sliding window between 192 and 352 samples for the 1-NN classifiers and between 288 and 352 samples for the SVM classifiers. In all cases, the accuracy of the classifiers drops significantly with a length smaller than 160 samples. Following these results, lengths of 224 and 192 samples were selected for the 1-NN classifiers with 24 channels and 8 channels, respectively; lengths of 288 and 320 samples were chosen for the SVM classifiers with 24 channels and 8 channels, respectively.

Then, the influence of the codebook size on the classification performance was evaluated. We fixed the step size to 4 and used the best window length obtained in the previous step for each classifier. The results in performance during the training process

are relatively stable with a codebook size between 50 and 1500 words for the 1-NN classifiers and between 500 and 3000 words for both SVMs. It is important to reduce computational time and memory usage while having good discrimination, thus the best values for codebook size were 500 words for both 1-NNs, and 2000 and 2500 words for the SVM classifiers with 8 channels and 24 channels, respectively.

Finally, the last parameter to evaluate was the step size of the sliding windows. In this case, the best values for the windows length and codebook size were used to test different step size from one to nine samples. The evaluation showed stable results for most values, however, the best results were obtained with step size of four and six samples for the 24-channel and 8-channels 1-NN, respectively, and five samples for both SVMs. Therefore, the best BoW parameters were used to build the codebook, which was universal for both training and testing of each classifier.

In addition, an example of cluster centers obtained by *k*-means is presented in Fig. 7. The image presents the DWT's approximation coefficients derived from the sliding window process. In this set of samples, we can observe distinctive



**Fig. 7** Samples of cluster centers constructed by *k*-means on the zero-mean and unit-variance fNIRS using the DWT coefficients.

**Table 1** Performance of the K-NN ( $K = 1$ ) and SVM (linear) classifiers using the best results from parameter tuning of the BoW representation. The evaluation also includes two sets of channels, all 24 channels (Ch) and 8 channels, from the RoI. "Low" refers to pain threshold, while "high" refers to pain tolerance.

Classifier	Ch	Step Size	Segment Length	Codebook Size	Class	Values (%)		
						Accuracy	Sensitivity	Specificity
K-NN	24	4	224	500	Low-cold	95.00	89.06	98.20
					Low-heat	89.66	92.86	96.57
					High-cold	93.22	98.21	97.65
					High-heat	90.48	89.06	96.47
					Average	92.08	92.30	97.22
SVM	24	5	288	2500	Low-cold	90.00	84.38	96.49
					Low-heat	94.55	92.86	98.24
					High-cold	93.10	96.43	97.63
					High-heat	88.06	92.19	95.24
					Average	91.25	91.46	96.90
K-NN	8	6	192	500	Low-cold	92.70	85.94	97.44
					Low-heat	94.67	95.24	98.82
					High-cold	89.39	95.24	96.33
					High-heat	89.69	90.63	96.04
					Average	91.53	91.76	97.01
SVM	8	5	320	2000	Low-cold	90.63	90.63	96.39
					Low-heat	92.73	91.07	97.66
					High-cold	91.38	94.64	97.06
					High-heat	88.89	87.50	95.86
					Average	90.83	90.96	96.74

patterns, i.e., increasing, decreasing, decreasing, and then increasing. Each of these centroids symbolizes the most representative value from a group of fNIRS segments with distinctive pattern similarities.

### 3.3 Classification of fNIRS Signals in Response to Thermal Pain

The main goal of the classification task is to design a classifier that can make predictions using new observations. As mentioned before, we have four types of observations (classes); these observations refer to temperature level (hot and cold) and pain intensity (low and high). Based on these observations we trained our classifier to predict new observations to be either low-cold pain, low-heat pain, high-cold pain, or high-heat pain. With this in mind, we used the K-NN (with  $K = 1$ , 1-NN) and the linear SVM algorithms with 24 channels and 8 channels (fNIRS channels) to design four classifiers able to estimate the type of pain and its intensity by using the BoW representation. Using the best BoW parameters for each classifier, we tested each classification model with unseen subjects ( $n = 5$ ), the results were averaged and summarized in Table 1.

The overall performance of each classifier showed good generalization results with both channel sets. The best results were obtained by the 1-NN classifier with 92.08% and 91.53% using the 24 channels and 8 channels, respectively. On the other hand,

the SVM classification models showed a slightly lower performance compared to the 1-NN classifiers, 91.25% using 24 channels, and 90.83% using the 8-channel set. In addition, we also evaluated different NNs with  $K = 3, 5, 7, 9$ ; however, in both channel sets, the 1-NN showed the best results.

## 4 Discussions

In this study, we demonstrated that fNIRS and machine learning are powerful tools to identify thermal pain and its intensity in a research setting. We used the BoW representation of fNIRS signals and compared two state-of-the-art classifiers (K-NN and SVM) to predict the response of new fNIRS within four categories of painful stimulation (low-cold, low-heat, high-cold, and high-hot) using two sets of channels (24 channels and 8 channels). The classification models exhibited good classification accuracy (90.83% to 92.08%) and demonstrated the possibility to design classifier models that can differentiate multiple modalities of pain and their intensity. This study supports the idea of developing an objective assessment of pain for nonverbal patients.

### 4.1 Building the Classifiers

The BoW representation showed its potential to characterize fNIRS signals, however, parameter tuning of such representation is important to obtain a trade-off between discrimination

and generalization. While the step size of the sliding window showed stable results with different numbers of steps, the segment length and the codebook size presented more variability. Local information was extracted by continuously sliding a window along an fNIRS signal and BoW allows incorporating this information that appears at different temporal locations and with different scales in an efficient way. If the window segment is short, the local information loses meaningfulness due to the lack of data points for feature extraction and a large number of local segments will be extracted from each fNIRS signal, which increases the computational complexity due to the large number of segments to cluster. However, if the window segment is large, only a few local segments are obtained, which restricts the discriminative ability of the BoW representation due to the limited number of samples. On the other hand, the codebook size showed good results with medium-size codebooks, however, it varied for K-NN and SVM classifiers. In previous works on classification in biomedical signals, such as EEG and ECG<sup>30,34,36</sup> using BoW, the optimal codebook size has ranged from 25 to 8024 words, and this diversity is mainly due to differences between the nature and data type of each study. For our particular study, a compact codebook (<500 words) produced limited discriminative results because segments with a different profile can be assigned to the same cluster, whereas a large codebook (>2000 words) increased computational time and memory usage. Therefore, obtaining the most suitable values of BoW parameters for each specific data type produces better results while being more computationally efficient.

The generalization capability between the classification models for the discrimination of pain showed that the K-NN classifier performed slightly better than SVM on both channel sets. This can be explained due to the fact that the K-NN classifier was designed with a smaller codebook (500 words) compared to the SVM classifier (2000 to 2500 words). Thus, increasing the number of words amplifies the intraclass variability, which decreases classification accuracy and also increases computational time and memory usage. We also found that unlike SVM, K-NN was much more adept at classifying low-cold (threshold of cold) and high-heat (tolerance of heat) on both channel sets, which might be the reason why the accuracy rate for SVM was slightly lower compared to K-NN. The difference between these two modalities (low-cold and high-heat) of pain might also suggest that distinctive patterns from these two classes are more inconsistent among subjects and thus more difficult to separate in a high dimensional space using a linear kernel. As mentioned before, we also compared different values of NNs ( $K = 3, 5, 7, 9$ ), however, the performance dropped; the risk of including NNs that belong to different classes increases since the K-NN algorithm uses a majority voting scheme.

Multiple cortical areas were monitored using a 24-channel head probe, because of that, we were interested to know if the hemodynamic response in a specific RoI could provide better classification accuracy compared to the whole probe. For that reason, we selected the somatosensory area as RoI and trained the classifiers with data only from this cortical area using (eight channels). The reasoning behind this comparison was that the somatosensory area, in particular, the primary (S1) and secondary (S2) somatosensory cortex, has shown a fundamental role in pain processing.<sup>8,12,18</sup> The results showed that the accuracy of RoI-based classifiers was slightly lower than whole-probe classifiers (refer to Table 1). For instance, for the K-NN classifier,

the difference between RoI-based and whole-probe approach was a loss of  $\sim 0.55\%$  in accuracy, whereas for the SVM classifiers, the accuracy dropped  $0.42\%$ . A clear advantage of using an RoI-based approach is the need for less channels, which decreases the computational cost to build the BoW representation and the classification models (K-NN and SVM), since the RoI-based dataset is built with one third of the original dataset. However, the RoI-based approach did not perform better than the whole-probe approach in either case. This finding is consistent with Brown et al.,<sup>18</sup> the authors found that using a whole-brain approach produced better accuracy to classify pain and no-pain signals using a linear SVM compared to an RoI-based approach. These findings confirm that despite the fact that the somatosensory region plays a fundamental role in pain processing, there are still more cortical regions than are implicated in sensory aspects of pain perception (e.g., insular cortex, anterior cingulate cortex, prefrontal cortex), which make nociceptive processing a large distributed brain network, referred as the "pain matrix."<sup>44</sup> Therefore, it is also important to monitor different cortical regions that may be involved in the particular pain experience, the interaction between different cortical areas could tell more about the brain functioning as a system rather than looking at separate areas working independently.<sup>13</sup>

## 4.2 Pain Assessment

The definition of pain is sometimes broad and controversial and is mainly regulated by the perceived intensity of the painful sensation. In this study, we differentiated the intensity of a painful stimulus as pain threshold (low pain) and pain tolerance (high pain). It is believed that all humans have a similar threshold for pain but different pain tolerance.<sup>45</sup> For instance, heat is perceived as noxious at thermal doses between  $40^{\circ}\text{C}$  and  $45^{\circ}\text{C}$ , same temperature at which tissue damage occurs, while the response to pain intensities differs from person to person and is influenced by factors, such as, gender, age, weight, etc. In addition, in some circumstances, the tolerance to pain of a person can change after a stressful experience (e.g., death of a loved one, self-injury). In a study by McCoy et al.,<sup>46</sup> it was found that participants with previous experience of self-injury had a significantly higher pain tolerance but the same pain threshold compared to normal controls; it was hypothesized that self-injury increases the desensitization to fear and pain.

An advantage of obtaining the intensity (low and high) of pain in a patient is to provide a personalized medical treatment. It is well reported that pain judgement is affected by factors, such as age, gender, weight, cultural background, etc. Medical treatments that work well for some patients may not be adequate for others due to physical differences, therefore tailoring pain control to the individual needs of each patient is not only important for nonverbal patients but also for the whole population. In many cases, medical practitioners prescribe analgesics based only on general information about medical treatments and not on the individual patient's characteristics. For this reason, knowing the individual pain sensation will reduce the risks of harmful side effects due to overdose and cut medical treatment costs compared with traditional trial-and-error treatments.

Thermal testing was kept as uniform as possible among all the participants, however, possible differences in data acquisition could occur. In order to minimize any differences between observations and thermal acquisition, we utilized thermode-induced stimulation. In contrast with water-based methods,

such as cold pressor test and the hot water immersion test,<sup>47</sup> the thermode induces heat or cold from the baseline of 32°C at a constant rate. One major difference between the two methods is the size of the area of application; while water-based methods require one hand to be submerged into cold/hot water, the thermode can be used on any area of the skin with a small surface (9 cm<sup>2</sup>). Thermode-induced stimulus safely activates nociceptive peripheral afferents and induces a moderate to intense painful sensation. One important feature of thermode-based stimulation is its reproducibility, which reduces the intrasubject variability of the measurements and increases reliability of the thermal experiment. Thermode-induced stimulation is less time consuming, and it can be used routinely and repeatedly in patients, which makes it ideal for clinical use. In addition, thermode-based stimulation safely activates A-delta (A $\delta$ ) and C fibre nociceptors, which are known to transmit sensory information to the spinal cord and the brain. Nociceptors are specialized sensory nerve endings that warn us of potential tissue damage by identifying harmful temperature, pressure, and chemicals; nociceptors convert these stimuli into electrical signals that are conveyed through the central nervous system to higher cortical regions in the brain. A $\delta$  and C fibre nociceptors carry noxious sensory information and they have different functions while transmitting sensory information. Fine myelinated A $\delta$  fibers respond to mechanical and cold stimuli, transporting rapid and sharp pain; unmyelinated C fibers respond to warm detection and heat pain, transmits slow, burning pain, responding to chemical, mechanical and thermal stimuli.<sup>22</sup>

### 4.3 Limitations

We acknowledge that one potential issue was the lack of control for any skin blood flow contributions and intracerebral hemodynamics to the fNIRS signals. Recent studies have highlighted the issue that fNIRS signals encompass not only hemodynamic fluctuations due to neurovascular coupling but also due to skin blood flow and task-related systemic activity of cortex.<sup>23,28</sup> On the one hand, extracerebral hemodynamics is originated at shallow tissue (e.g., scalp), the area in which NIR light travels before reaching the cerebral cortex, and thus, contributions to the observed fNIRS signals of HbO and Hb are expected. Therefore, it is important to account for probable influence of scalp blood flow, which may or may not be temporally consistent with cortical signals. In our study, despite possible contributions of extracerebral confounders, the observed activation was measured in areas known to be involved in pain processing, which suggests that the fNIRS signals reflect mainly localized cortical vascular dynamics. On the other hand, intracerebral hemodynamics is caused by systemic physiological interference that may resemble true task-related cortical activities. These physiological variables are mainly associated with cardiac pulsations, respiration, arterial blood pressure, and activity in the autonomic nervous system.<sup>23</sup> In our case, the peak amplitudes for the HbO response to thermal stimuli were substantially distinguishable relative to the baseline signal at resting state. In both cases, we tried to reduce these confounders (extra- and intracerebral) by (1) using PCA on the baseline data to identify those PC that account for most of the variance while in resting state, and then reduce these PCs (potentially both, extra- and intracerebral confounders) in the stimulus data. (2) Using wavelet analysis in frequency bands that are only related to the expected period (Hz) of response and therefore avoiding those bands related to heartbeat (~1 Hz), respiration (~0.3 Hz),

or blood pressure (~0.1 Hz); however, very-low frequency oscillations (VLFO) associated with various brain stem nuclei in the sympathetic tone of cerebral arterioles (0.02 to 0.08 Hz)<sup>48</sup> might overlap with the content of the stimulus response, since the period of our block-design stimulation (~30 s) corresponds to this band. Finally, using both methods might not be enough to reduce such sources of hemodynamic interference and other methods, such as using short-separation channels to reveal scalp blood flow, or employing additional instruments to measure physiological signals would be desirable to regress out of these global signals.

### 4.4 Future Work and Final Remarks

The results of this study expand earlier findings from previous research of pain assessment using neuroimaging methods.<sup>18,19,42</sup> Direct comparisons are difficult because of the use of different experimental conditions, neuroimaging technique (e.g., fMRI), sampled population, data representation (using BoW), validation method, and classification models. Therefore, the contributions of this study can be summarized as follows: (1) presenting the classification of two types of thermal stimuli (cold and heat) and the discrimination between corresponding low (threshold) and high (tolerance) pain for each thermal stimulus, (2) showing that the pain tolerance and pain threshold test from the QST could be used to obtain pain information from nonverbal patients, (3) comparing two classification models (K-NN and SVM) to classify four different modalities of pain, and (4) investigating whether feature extraction from the somatosensory area could produce better classification accuracy compared to whole-probe approach.

The development of a bedside monitor for the diagnosis of pain is yet to be fully established. To expand such method into clinical settings requires validation on new datasets (generalization), include more participants of different populations (e.g., age, background), use multiple types of noxious stimulations (e.g., mechanical, chemical), stimulate different parts of the body (robust to testing location), and include models to study chronic pain disorders (e.g., back pain, neuropathic pain, fibromyalgia). In addition, the use of more learning models should be evaluated, such as decision trees, linear discriminant analysis (LDA), neural networks (NN), or higher-order SVM kernels (e.g., radial basis function, polynomial).

fNIRS has demonstrated to be a method that has potential for the assessment of pain. fNIRS is a technique capable of identifying cortical hemodynamic changes in response to chemical, temperature, and pressure noxious stimuli. In addition, our study demonstrates that the use of fNIRS in combination with machine learning techniques is a powerful tool for the assessment of pain in experimental settings. fNIRS possesses advantages over PET or fMRI to be used in more realistic clinical settings, e.g., less exposure to ionizing radiation, safe to use over long periods and many times, less expensive, easy to use, and small size. Finally, our findings advance knowledge in pain assessment using neuroimaging as a method of diagnosis and represent a step closer to developing a physiologically based diagnosis of human pain that would benefit vulnerable population who cannot self-report pain.

### Disclosures

The authors have no relevant financial interests in the manuscript and no other potential conflicts of interest to disclose.

## Acknowledgments

This work was partially supported by the Human-Centred Research Centre (FFR) and 3D Global Biotech Inc. (K-L.O). The authors thank Lancelot Caddy for proofreading this paper.

## References

1. C. L. von Baeyer, "Childrens self-report of pain intensity: what we know, where we are headed," *Pain Res. Manage.* **14**(1), 39–45 (2009).
2. K. Herr et al., "Pain assessment in the patient unable to self-report: position statement with clinical practice recommendations," *Pain Manage. Nurs.* **12**(4), 230–250 (2011).
3. M. S. Bates, W. T. Edwards, and K. O. Anderson, "Ethnocultural influences on variation in chronic pain perception," *Pain* **52**(1), 101–112 (1993).
4. R. R. Edwards et al., "Ethnic differences in pain tolerance: clinical implications in a chronic pain population," *Psychosom. Med.* **63**(2), 316–323 (2001).
5. I. Lund et al., "Gender differences in electrical pain threshold responses to transcutaneous electrical nerve stimulation (tens)," *Neurosci. Lett.* **375**(2), 75–80 (2005).
6. M. Ranger and C. Gélinas, "Innovating in pain assessment of the critically ill: exploring cerebral near-infrared spectroscopy as a bedside approach," *Pain Manage. Nurs.* **15**(2), 519–529 (2014).
7. M. Ferrari and V. Quaresima, "A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application," *Neuroimage* **63**(2), 921–935 (2012).
8. M. A. Yücel et al., "Specificity of hemodynamic brain responses to painful stimuli: a functional near-infrared spectroscopy study," *Sci. Rep.* **5**, 9469 (2015).
9. M. Ferrari, L. Mottola, and V. Quaresima, "Principles, techniques, and limitations of near infrared spectroscopy," *Can. J. Appl. Physiol.* **29**(4), 463–487 (2004).
10. M. Wolf, M. Ferrari, and V. Quaresima, "Progress of near-infrared spectroscopy and topography for brain and muscle clinical applications," *J. Biomed. Opt.* **12**(6), 062104 (2007).
11. F. Scholkmann et al., "A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology," *Neuroimage* **85**, 6–27 (2014).
12. A. V. Apkarian et al., "Human brain mechanisms of pain perception and regulation in health and disease," *Eur. J. Pain* **9**(4), 463–463 (2005).
13. R. Fernandez Rojas, X. Huang, and K.-L. Ou, "Region of interest detection and evaluation in functional near infrared spectroscopy," *J. Near Infrared Spectrosc.* **24**(4), 317–326 (2016).
14. R. Peyron, B. Laurent, and L. Garcia-Larrea, "Functional imaging of brain responses to pain. A review and meta-analysis (2000)," *Neurophysiol. Clin. Clin. Neurophysiol.* **30**(5), 263–288 (2000).
15. C.-H. Lee et al., "Analysis for distinctive activation patterns of pain and itchy in the human brain cortex measured using near infrared spectroscopy (NIRS)," *PLoS One* **8**(10), e75360 (2013).
16. M. Ranger et al., "A multidimensional approach to pain assessment in critically ill infants during a painful procedure," *Clin. J. Pain* **29**(7), 613–620 (2013).
17. R. Slater et al., "Cortical pain responses in human infants," *J. Neurosci.* **26**(14), 3662–3666 (2006).
18. J. E. Brown et al., "Towards a physiology-based measure of pain: patterns of human brain activity distinguish painful from non-painful thermal stimulation," *PLoS One* **6**(9), e24124 (2011).
19. T. D. Wager et al., "An fMRI-based neurologic signature of physical pain," *N. Engl. J. Med.* **368**(15), 1388–1397 (2013).
20. M. Gram et al., "Machine learning on encephalographic activity may predict opioid analgesia," *Eur. J. Pain* **19**(10), 1552–1561 (2015).
21. A. Pourshoghi, I. Zakeri, and K. Pourrezaei, "Application of functional data analysis in classification and clustering of functional near-infrared spectroscopy signal in response to noxious stimuli," *J. Biomed. Opt.* **21**(10), 101411 (2016).
22. R. Rolke et al., "Quantitative sensory testing in the German research network on neuropathic pain (DFNS): standardized protocol and reference values," *Pain* **123**(3), 231–243 (2006).
23. I. Tachtsidis and F. Scholkmann, "False positives and false negatives in functional near-infrared spectroscopy: issues, challenges, and the way forward," *Neurophotonics* **3**(3), 031405 (2016).
24. T. Yamamoto and T. Kato, "Paradoxical correlation between signal in functional magnetic resonance imaging and deoxygenated haemoglobin content in capillaries: a new theoretical explanation," *Phys. Med. Biol.* **47**(7), 1121–1141 (2002).
25. G. Strangman et al., "A quantitative comparison of simultaneous bold fMRI and NIRS recordings during functional brain activation," *Neuroimage* **17**(2), 719–731 (2002).
26. B. Molavi, G. Dumont, and B. Shadgan, "Motion artifact removal from muscle nir spectroscopy measurements," in *23rd Canadian Conf. on Electrical and Computer Engineering (CCECE)*, pp. 1–4, IEEE (2010).
27. A. Duncan et al., "Measurement of cranial optical path length as a function of age using phase resolved near infrared spectroscopy," *Pediatr. Res.* **39**(5), 889–894 (1996).
28. Y. Zhang et al., "Eigenvector-based spatial filtering for reduction of physiological interference in diffuse optical imaging," *J. Biomed. Opt.* **10**(1), 011014 (2005).
29. S. O'Hara and B. A. Draper, "Introduction to the bag of features paradigm for image classification and retrieval," arXiv preprint arXiv:1101.3354 (2011).
30. M. G. Baydogan, G. Runger, and E. Tuv, "A bag-of-features framework to classify time series," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(11), 2796–2802 (2013).
31. F. Briggs, R. Raich, and X. Z. Fern, "Audio classification of bird species: a statistical manifold approach," in *Ninth IEEE Int. Conf. on Data Mining (ICDM'09)*, pp. 51–60, IEEE (2009).
32. Z. Fu et al., "Music classification via the bag-of-features approach," *Pattern Recognit. Lett.* **32**(14), 1768–1777 (2011).
33. J.-J. Aucouturier, B. Defreville, and F. Pachet, "The bag-of-frames approach to audio pattern recognition: a sufficient model for urban soundscapes but not for polyphonic music," *J. Acoust. Soc. Am.* **122**(2), 881–891 (2007).
34. J. Wang et al., "Bag-of-words representation for biomedical time series classification," *Biomed. Signal Process. Control* **8**(6), 634–644 (2013).
35. J. Lin and Y. Li, "Finding structural similarity in time series data using bag-of-patterns representation," in *Int. Conf. on Scientific and Statistical Database Management*, pp. 461–477, Springer (2009).
36. J. Grabocka, M. Wistuba, and L. Schmidt-Thieme, "Scalable classification of repetitive time series through frequencies of local polynomials," *IEEE Trans. Knowl. Data Eng.* **27**(6), 1683–1695 (2015).
37. E. Kirilina et al., "Identifying and quantifying main components of physiological noise in functional near infrared spectroscopy on the prefrontal cortex," *Front. Hum. Neurosci.* **7**, 864 (2013).
38. Y.-G. Jiang et al., "Representations of keyword-based semantic concept detection: a comprehensive study," *IEEE Trans. Multimedia* **12**(1), 42–53 (2010).
39. F. Pereira, T. Mitchell, and M. Botvinick, "Machine learning classifiers and fMRI: a tutorial overview," *Neuroimage* **45**(1), S199–S209 (2009).
40. H. Ding et al., "Querying and mining of time series data: experimental comparison of representations and distance measures," in *Proc. of the VLDB Endowment*, Vol. **1**, pp. 1542–1552 (2008).
41. J. Lin, R. Khade, and Y. Li, "Rotation-invariant similarity in time series using bag-of-patterns representation," *J. Intell. Inf. Syst.* **39**(2), 287–315 (2012).
42. A. Marquand et al., "Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes," *Neuroimage* **49**(3), 2178–2189 (2010).
43. C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. Intell. Syst. Technol.* **2**(3), 27 (2011).
44. I. Tracey, "Imaging pain," *Br. J. Anaesth.* **101**(1), 32–39 (2008).
45. E. N. Marieb, *Human Anatomy & Physiology*, Benjamin-Cummings Publishing Company, Redwood City, California (1989).
46. K. McCoy, W. Fremouw, and D. W. McNeil, "Thresholds and tolerance of physical pain among young adults who self-injure," *Pain Res. Manage.* **15**(6), 371–377 (2010).
47. A. Streff et al., "Differential physiological effects during tonic painful hand immersion tests using hot and ice water," *Eur. J. Pain* **14**(3), 266–272 (2010).
48. M. Ter Laan et al., "Sympathetic regulation of cerebral blood flow in humans: a review," *Br. J. Anaesth.* **111**(3), 361–367 (2013).

Biographies for the authors are not available.